

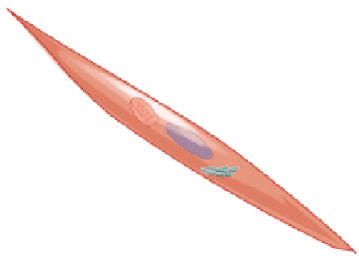
Computación Molecular Biológica

Dr. Edward Hinojosa Cárdenas

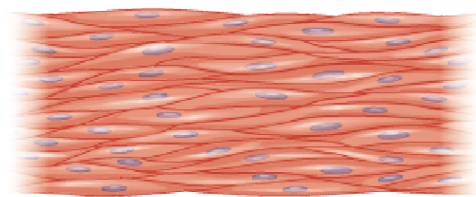
ehinojosa@unsa.edu.pe

Conceptos Iniciales

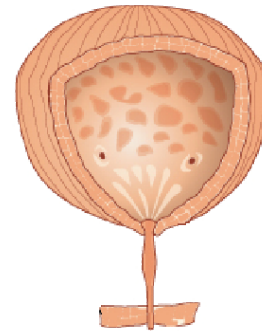
- Nuestros cuerpos están formado por un conjunto de órganos.
- Cada órgano está compuesto de un número de tejidos, y cada tejido es una colección de células similares que se agrupan para realizar funciones especializadas.



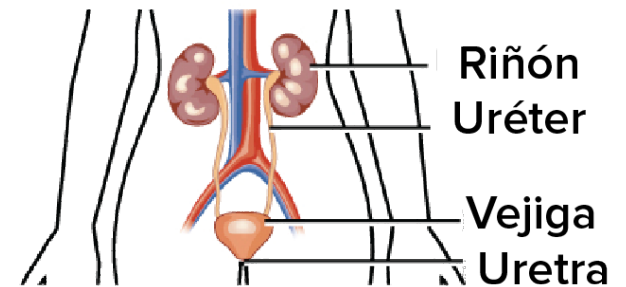
Célula muscular



Tejido muscular



Órgano (vejiga)

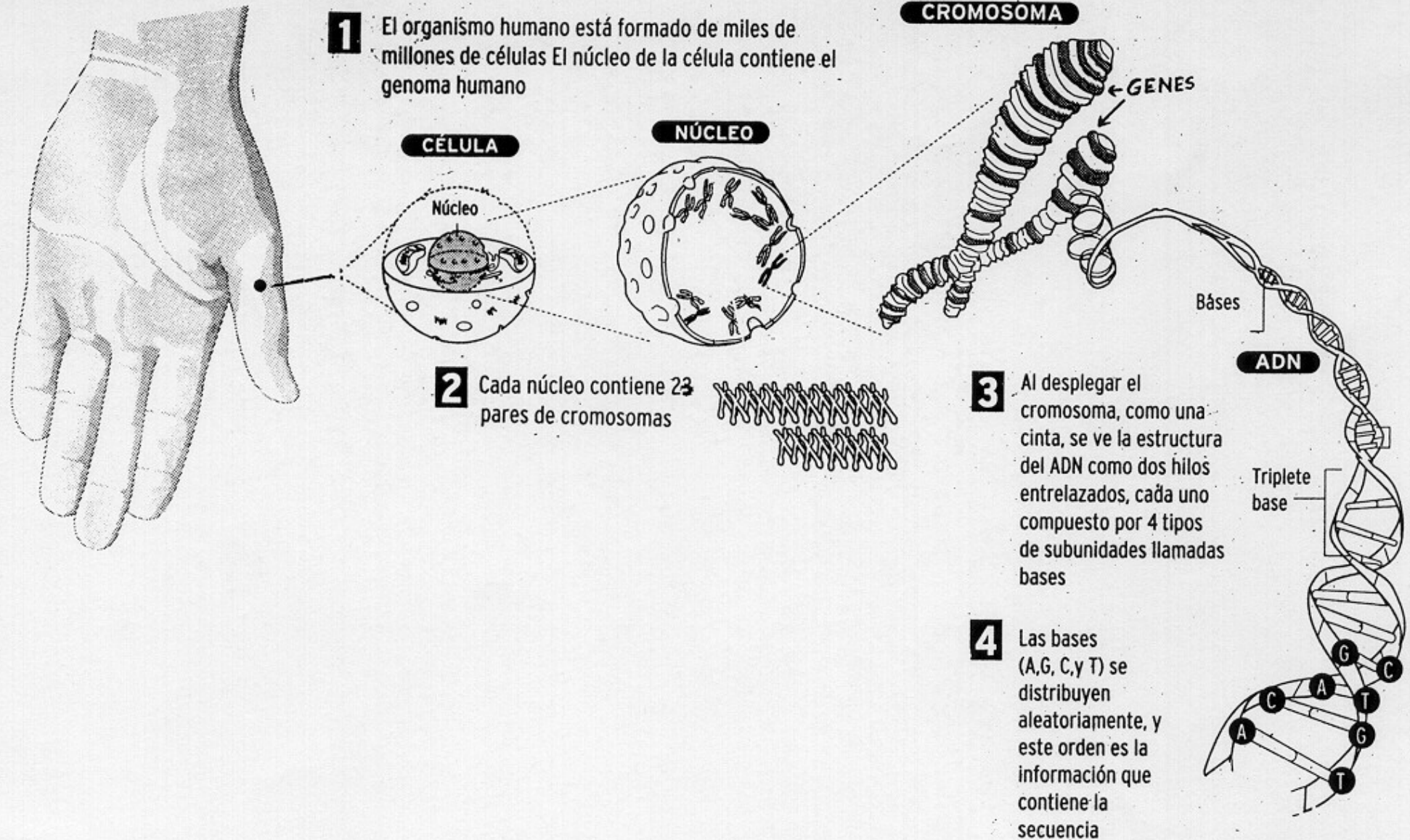


Sistema de órganos

Conceptos Iniciales



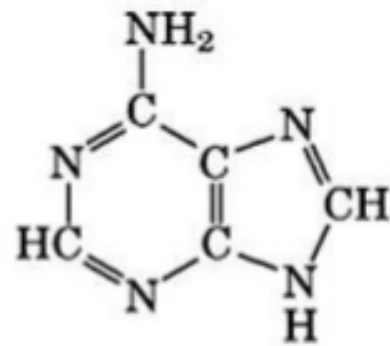
Conceptos Iniciales



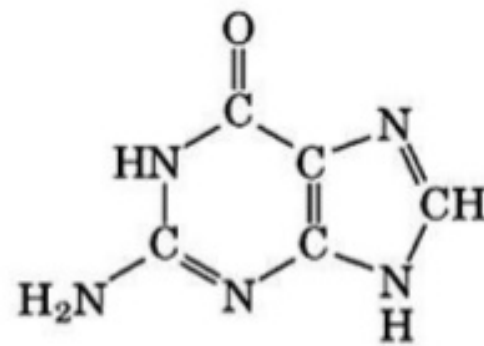
ADN

- La información genética en los seres vivos está contenida en las moléculas de ADN (ácido desoxirribonucleico).
- El ADN es una doble cadena de nucleótidos, los nucleótidos que forman el ADN sólo pueden ser cuatro:
 - A (adenina)
 - T (timina)
 - C (citosina)
 - G (guanina).

ADN

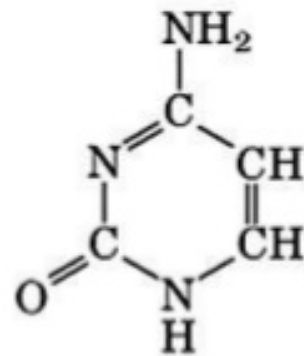


Adenina

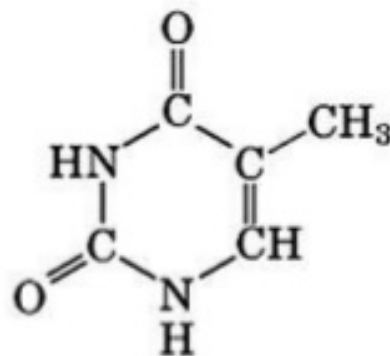


Guanina

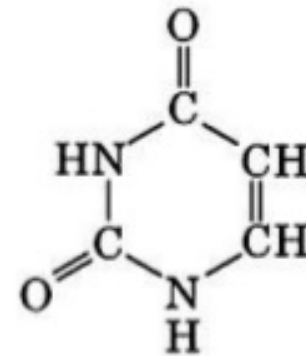
Purinas



Citosina



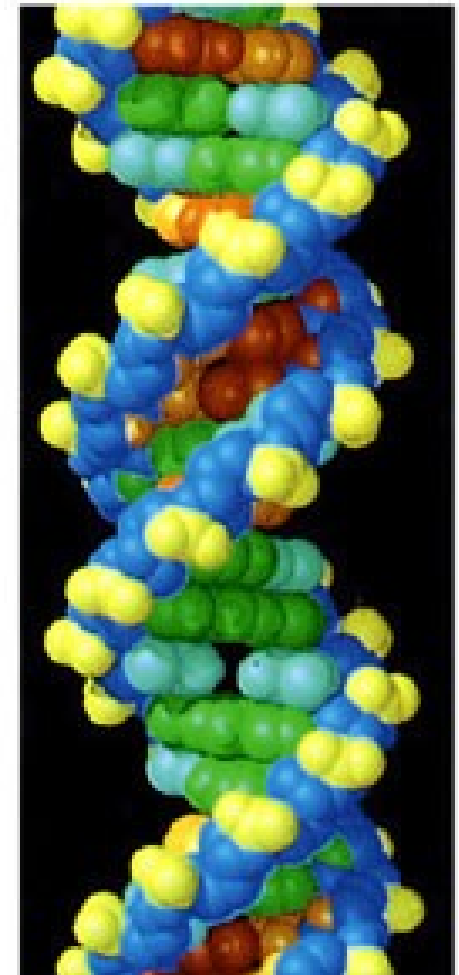
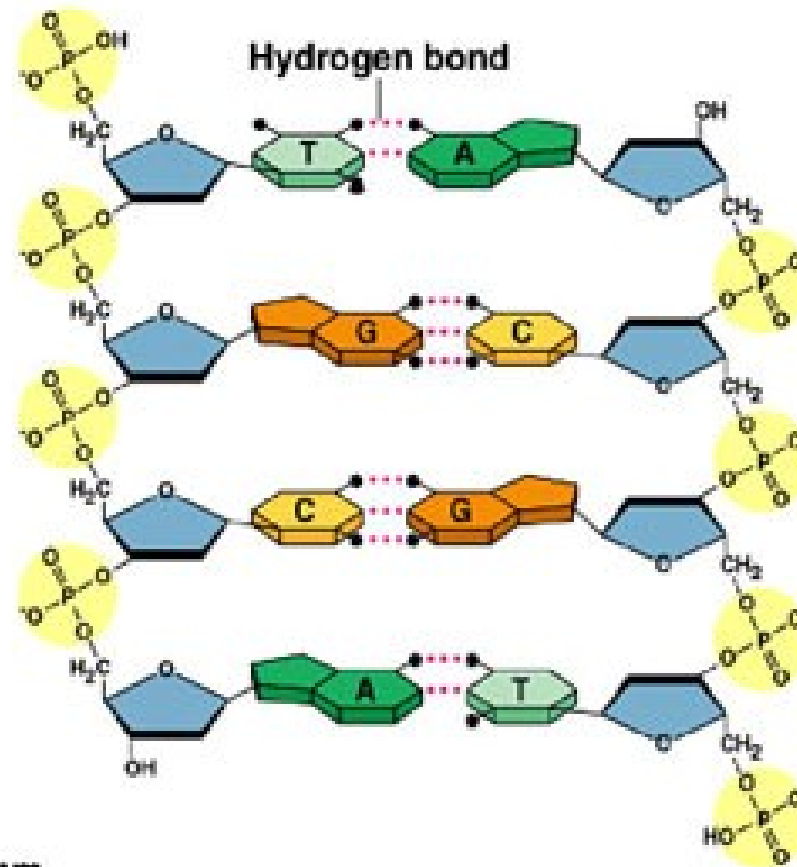
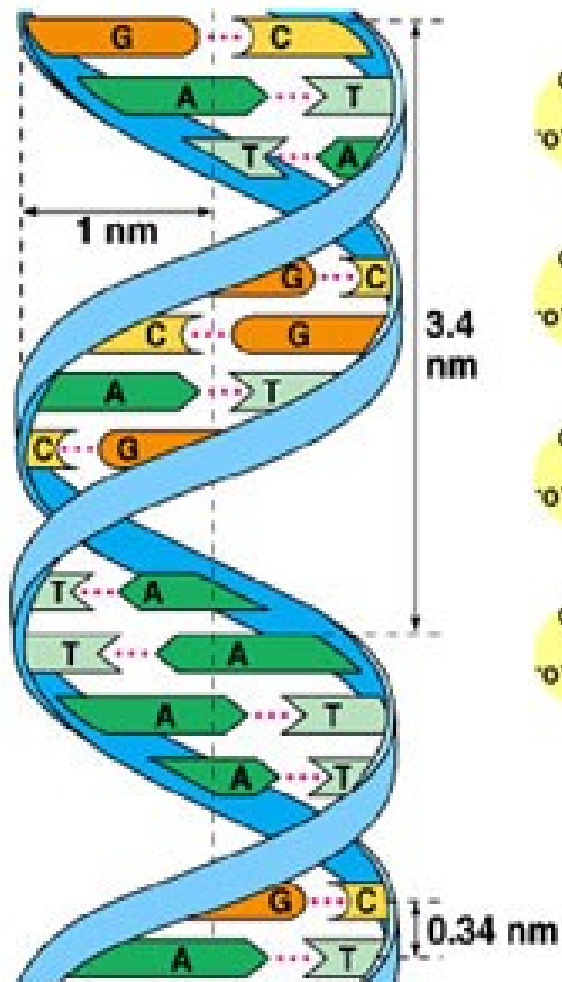
Timina (ADN)



Uracilo (ARN)

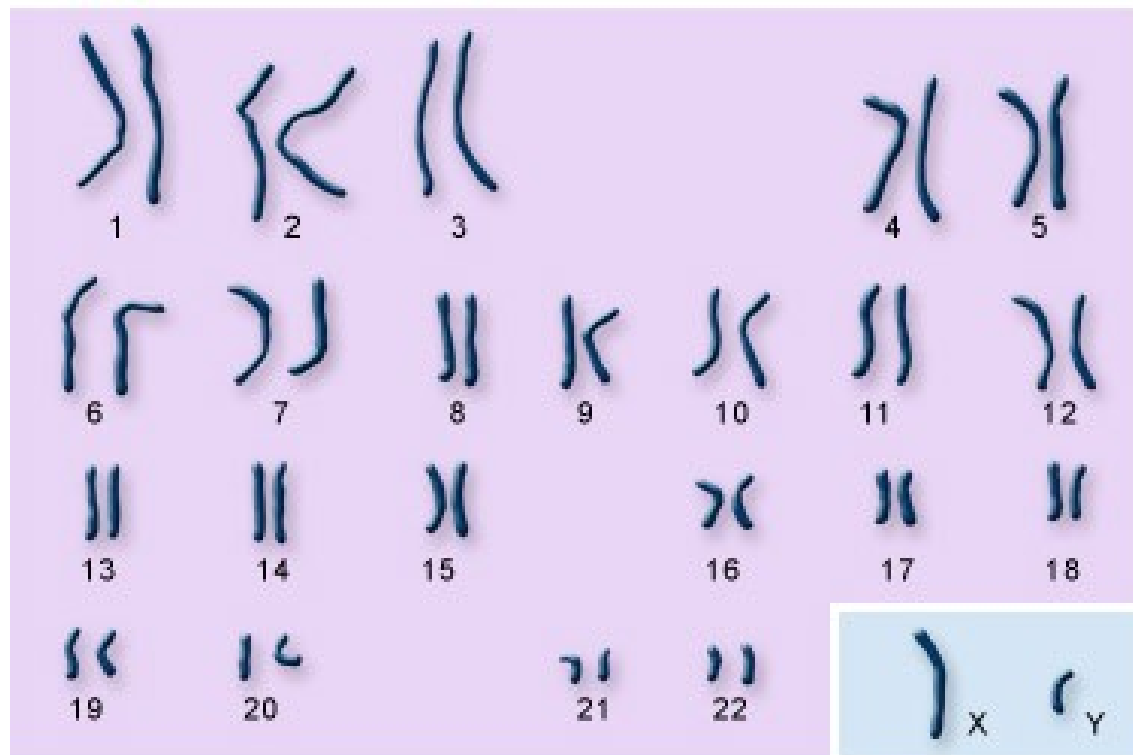
Pirimidina

ADN



ADN - Genoma

- El genoma humano es la secuencia de ADN de un ser humano. Está dividido en fragmentos que conforman los 23 pares de cromosomas distintos de la especie humano.



ARN

- Para que esta información pueda ser usada por la células debe transcribirse a una molécula de ARN (ácido ribonucleico).
- La molécula de ARN se copia fielmente a partir de la molécula de ADN en un proceso llamado transcripción. Existen diferencias químicas entre las moléculas que forman el ADN y el ARN, pero además el código difiere ya que la T del ADN es reemplazada por U (uracilo) en el ARN.

ARN

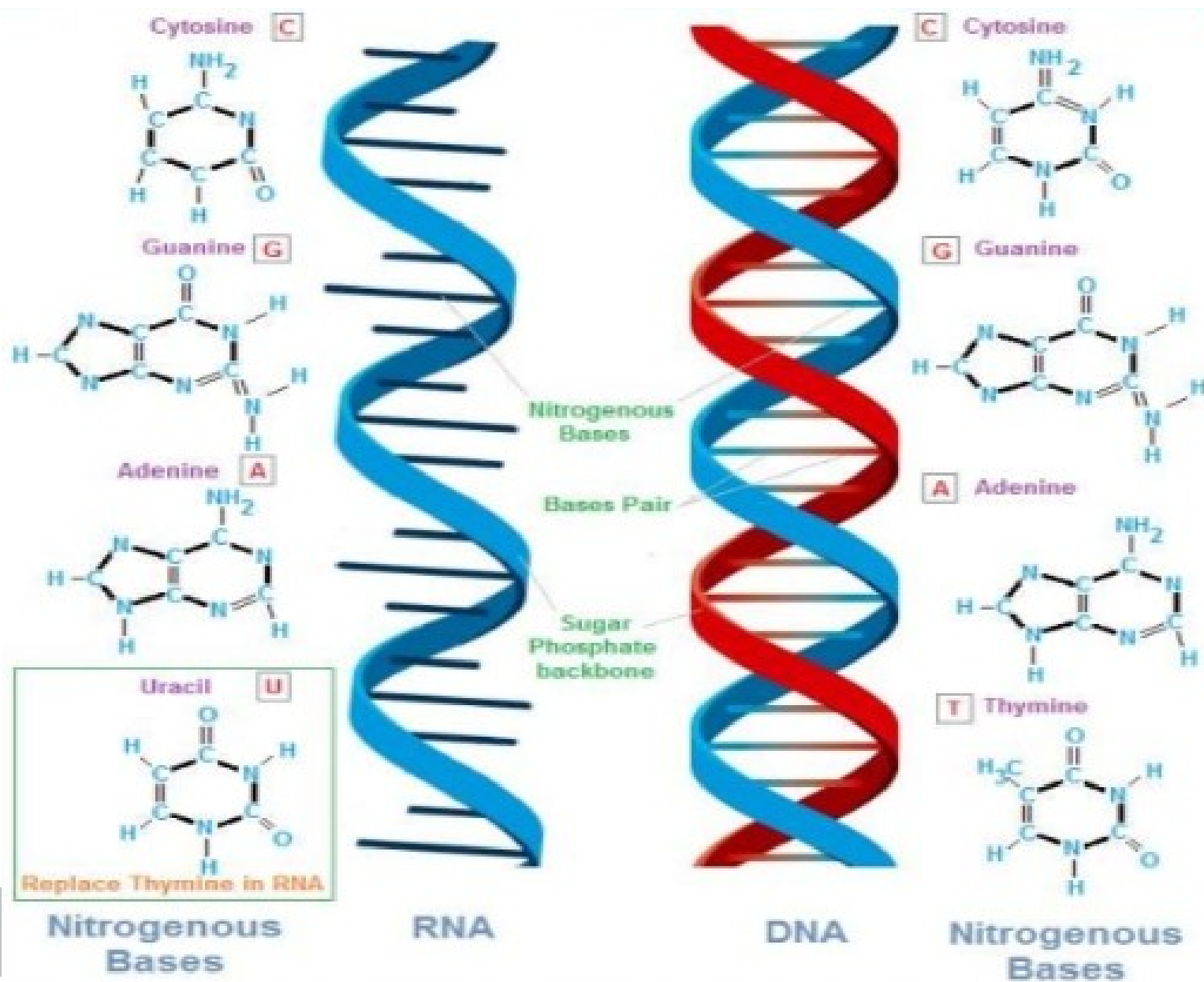
- Secuencia de ADN: Adenina, Timina, Guanina, Citosina.
- Secuencia de ARN: Adenina, Uracilo, Guanina, Citosina.

ACAGACAGATACAAT

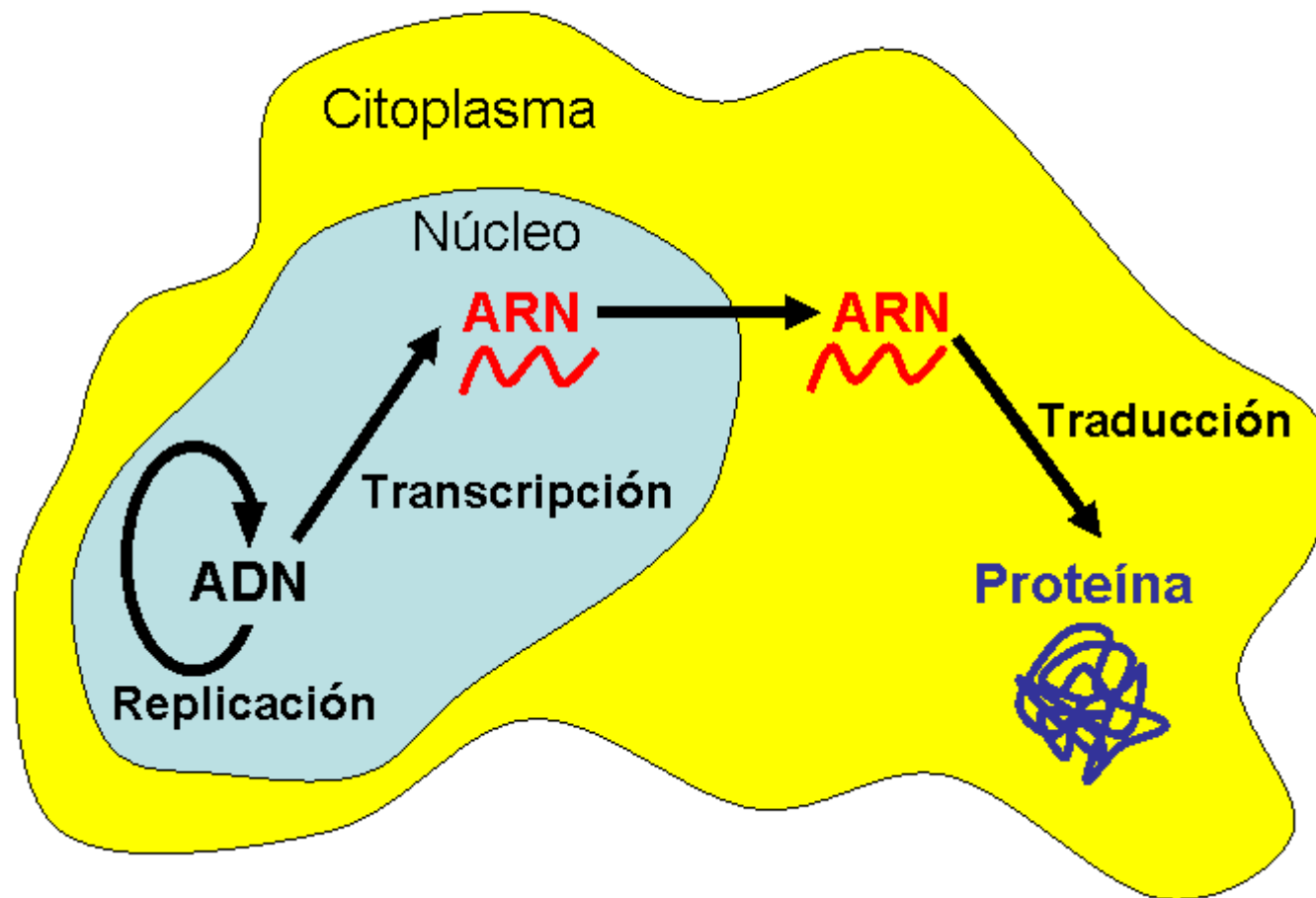
se transcribe a

ACAGACAGAUACAAU

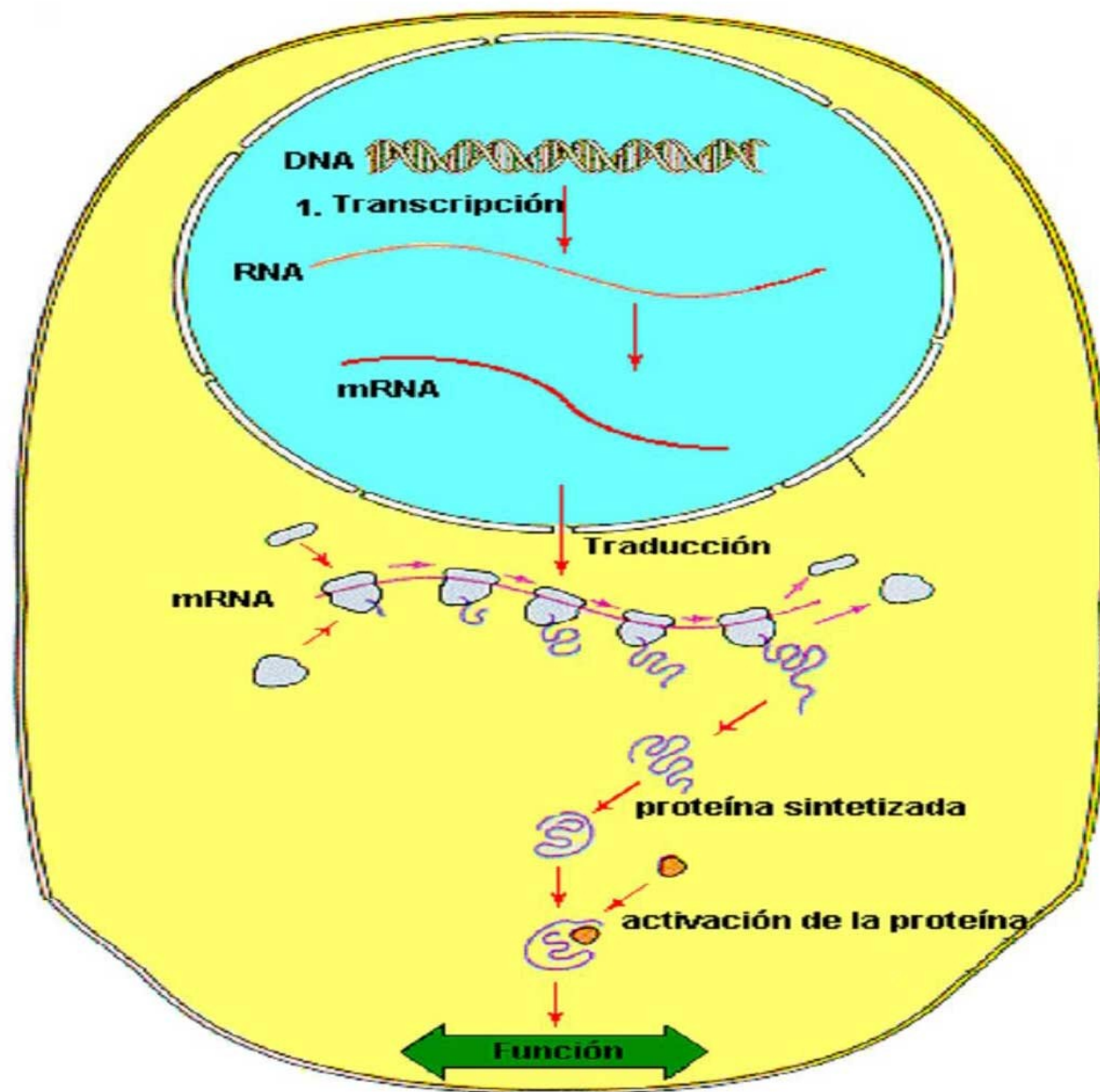
ARN



ARN

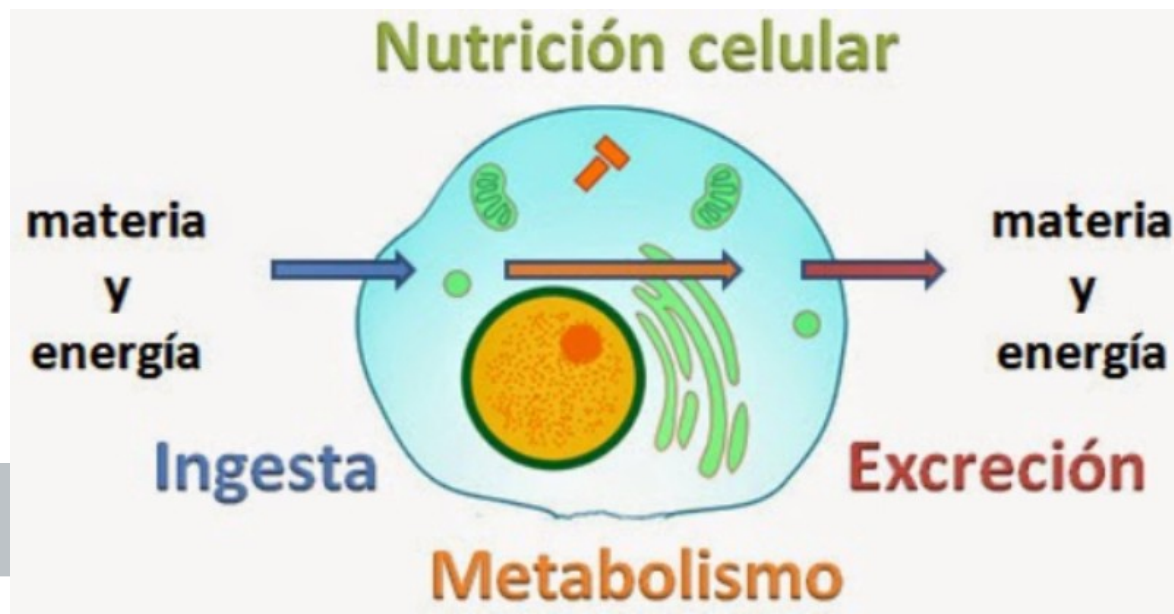


ARN



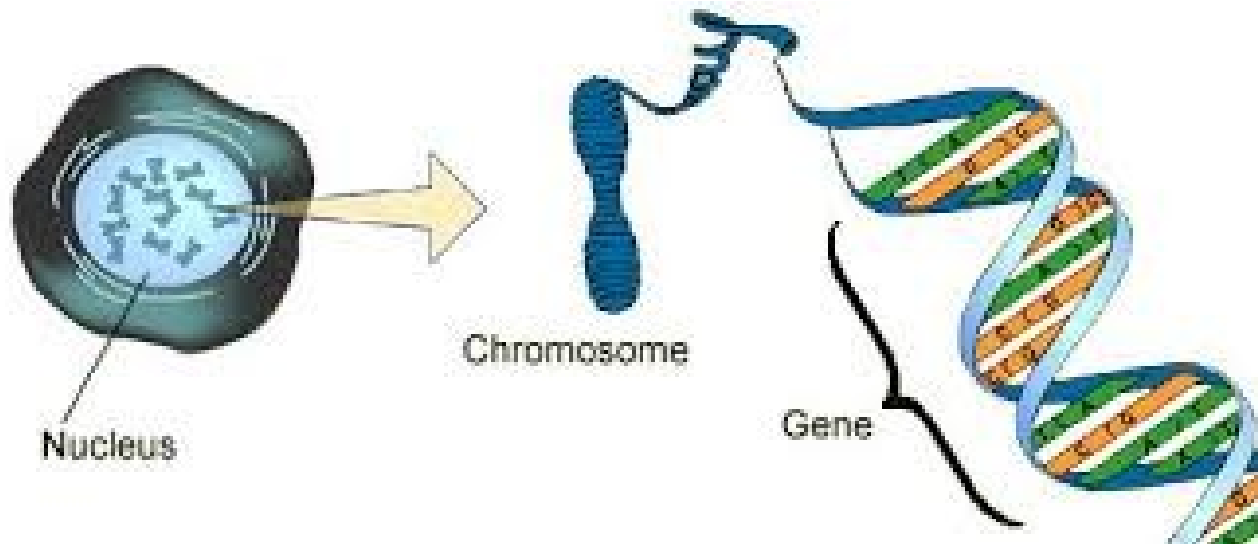
Células - Proteínas

- La célula individual es la unidad mínima de auto reproducción en todas las especies vivientes.
- Realizan dos tipos de funciones:
 - 1) Almacenar y transmitir la información genética para mantenerla viva de generación en generación
 - 2) Realizar reacciones químicas necesarias para mantener la vida.



Células - Proteínas

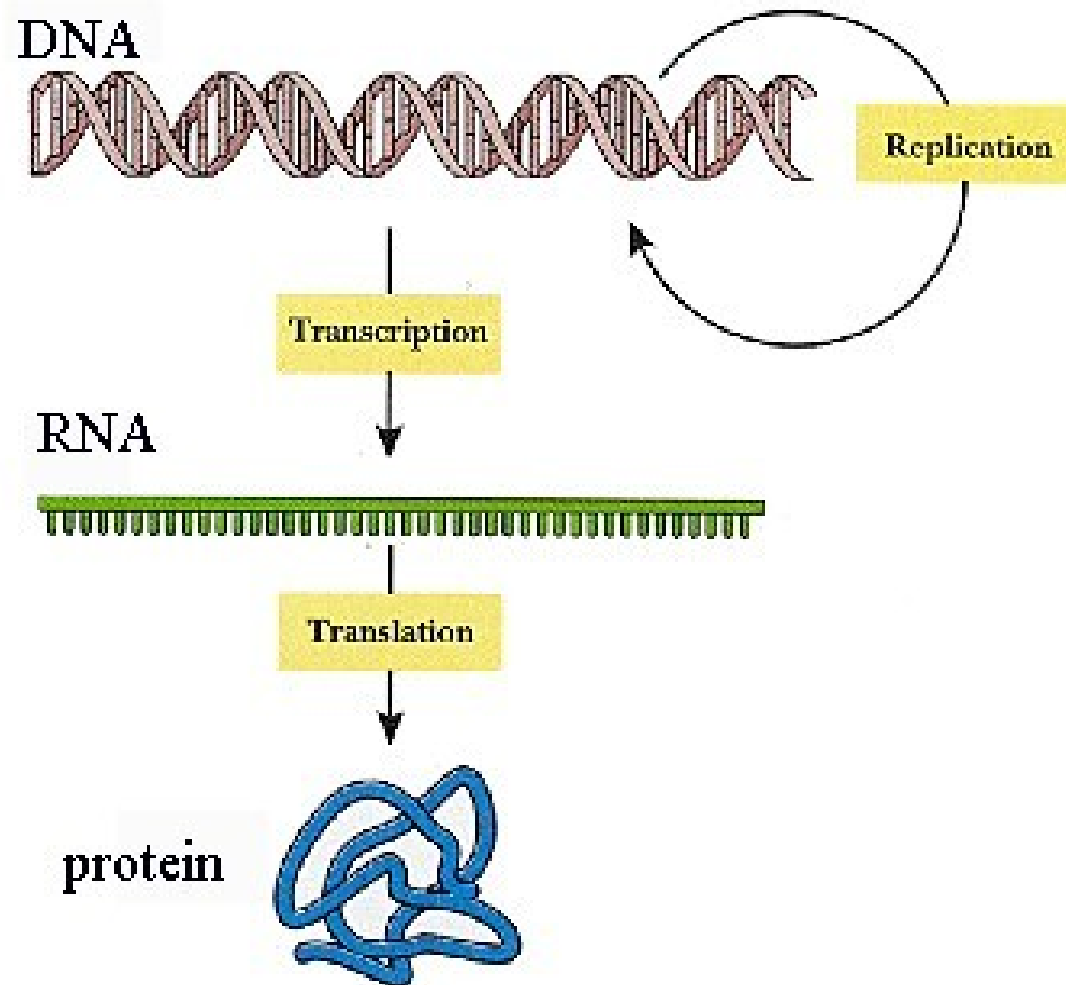
- Para la primera función, nuestras células almacena la información genética en forma de doble cadena DNA (Deoxyribonucleic acid) ADN (ácido desoxirribonucleico).



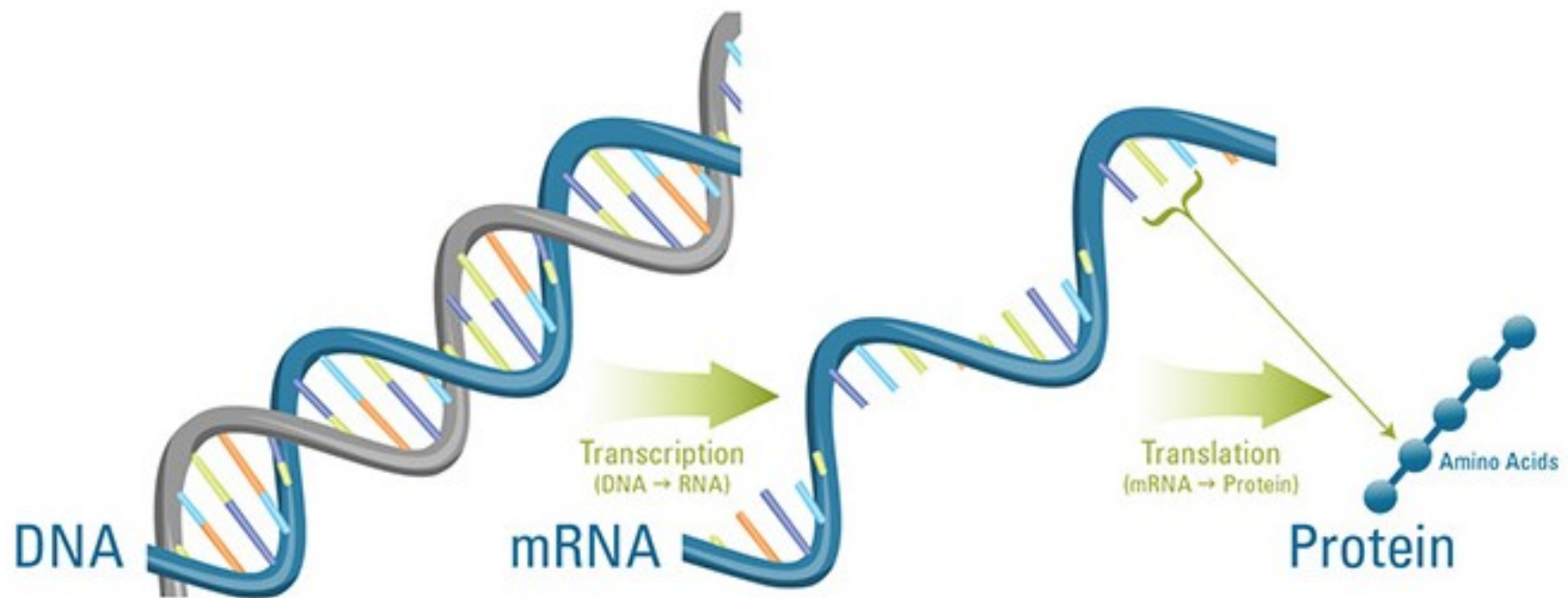
Células - Proteínas

- Para la segunda función, porciones del DNA llamadas de genes se transcriben en moléculas estrechamente relacionadas llamadas RNA (ribonucleic acid) ARN (ácido ribonucleico).
- RNA guía la síntesis de las moléculas de proteínas.

Células - Proteínas

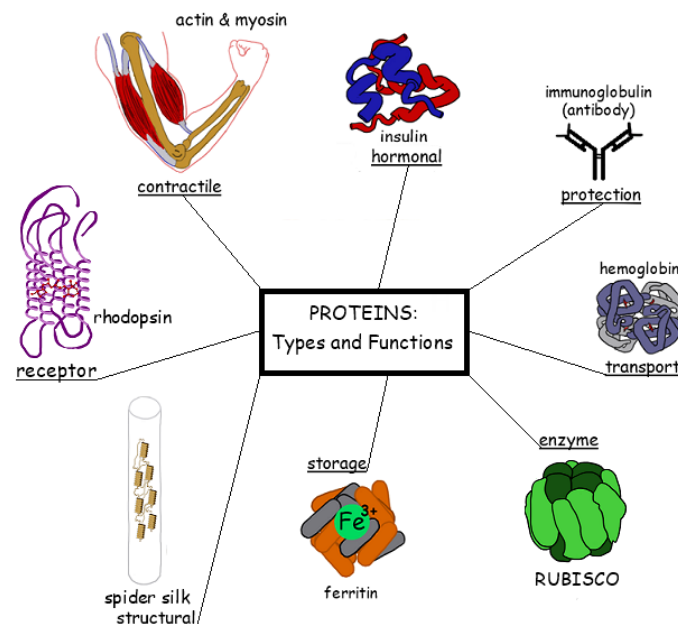


Células - Proteínas



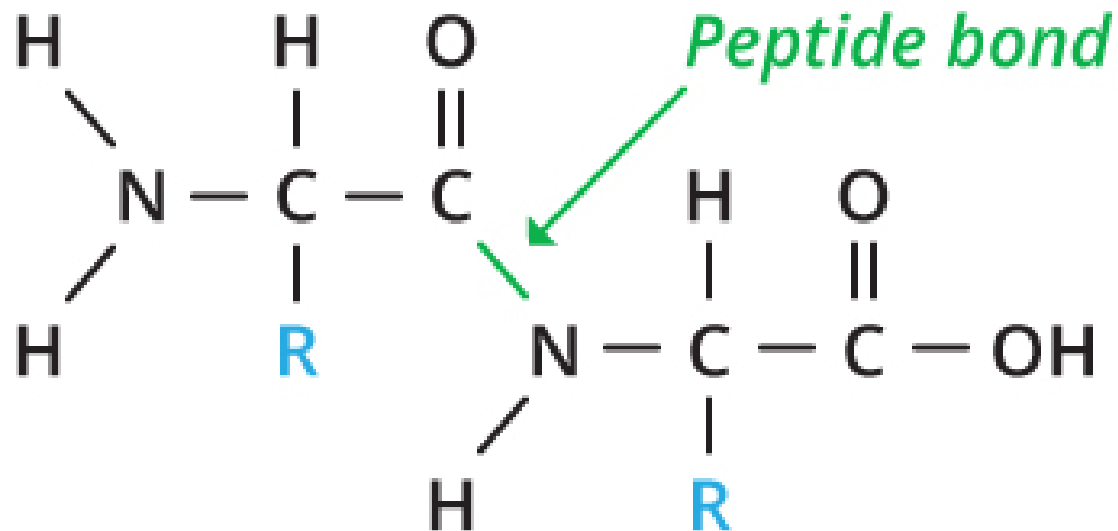
Proteínas

- Las proteínas constituyen la mayor parte de la masa seca de una célula. No solo son los bloques de construcción a partir de los cuales se construyen las células, sino que también ejecutan casi todas las funciones de la célula. Comprender las proteínas puede ayudarnos a comprender cómo funcionan nuestros cuerpos y otros procesos biológicos.



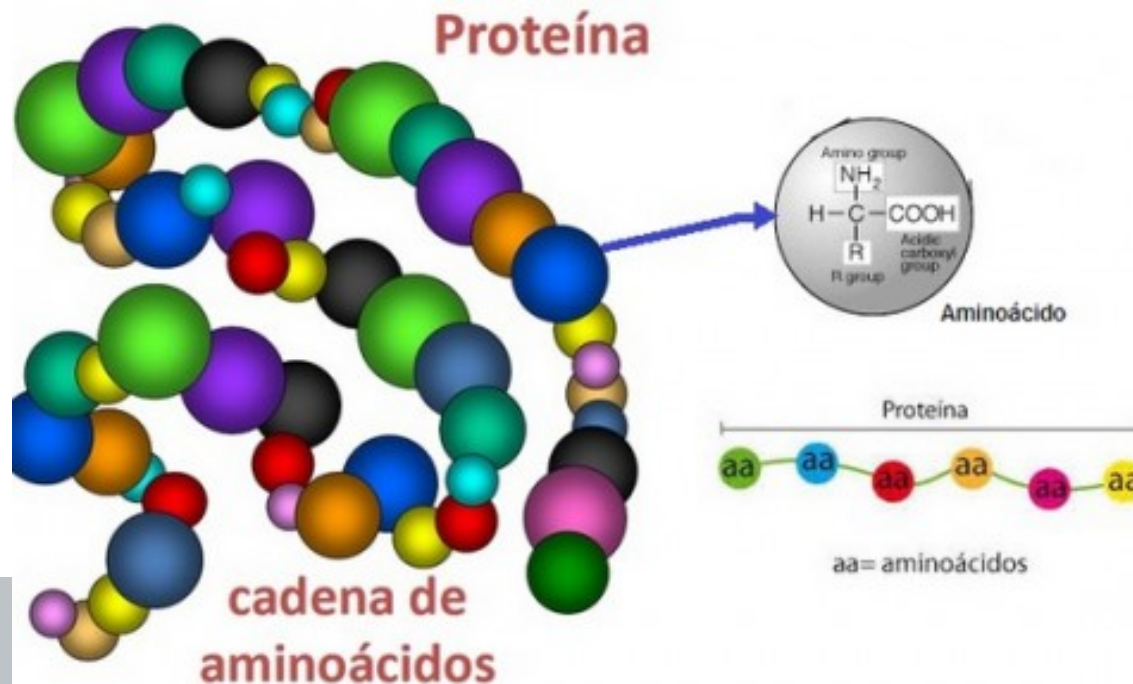
Proteínas

- Una proteína está compuesta de una cadena de aminoácidos, cada uno de los cuales se une a su vecino a través de un enlace peptídico covalente. Por lo tanto, las proteínas también se conocen como polipéptidos.



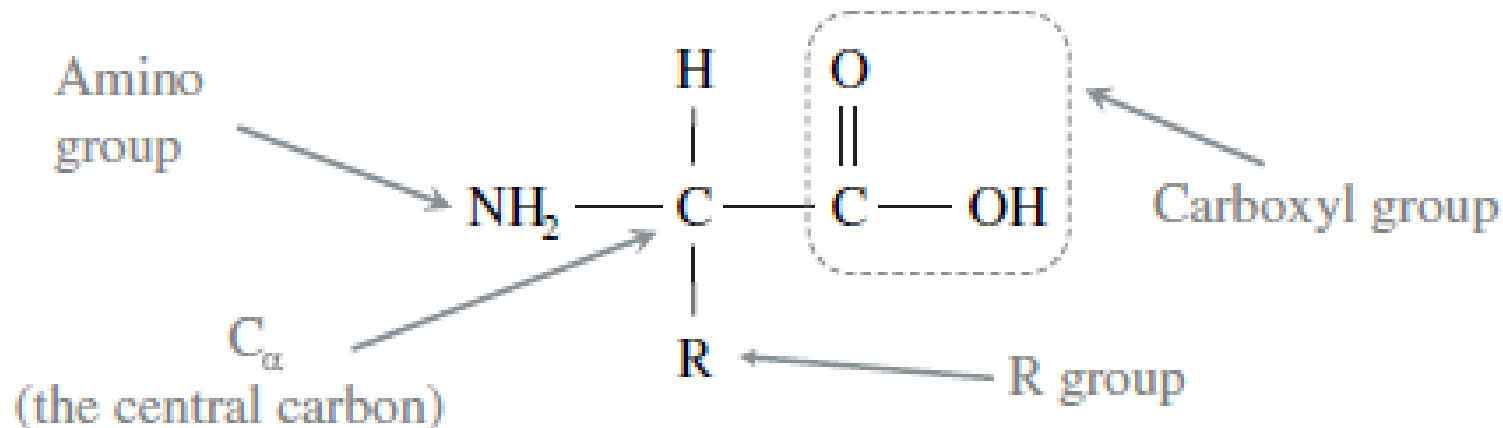
Proteínas

- Existen 20 tipos de aminoácidos y cada aminoácido tiene diferentes propiedades químicas. La longitud de una proteína está en el rango de 20 a más de 5000 aminoácidos. En promedio, una proteína contiene alrededor de 350 aminoácidos.



Aminoácidos

- Los aminoácidos son los componentes básicos de las proteínas. Cada aminoácido consiste en:
 - Grupo amino (grupo -NH_2)
 - Grupo Carboxílico (grupo -COOH)
 - Grupo R (El cual determina el tipo de un aminoácido)



Aminoácidos

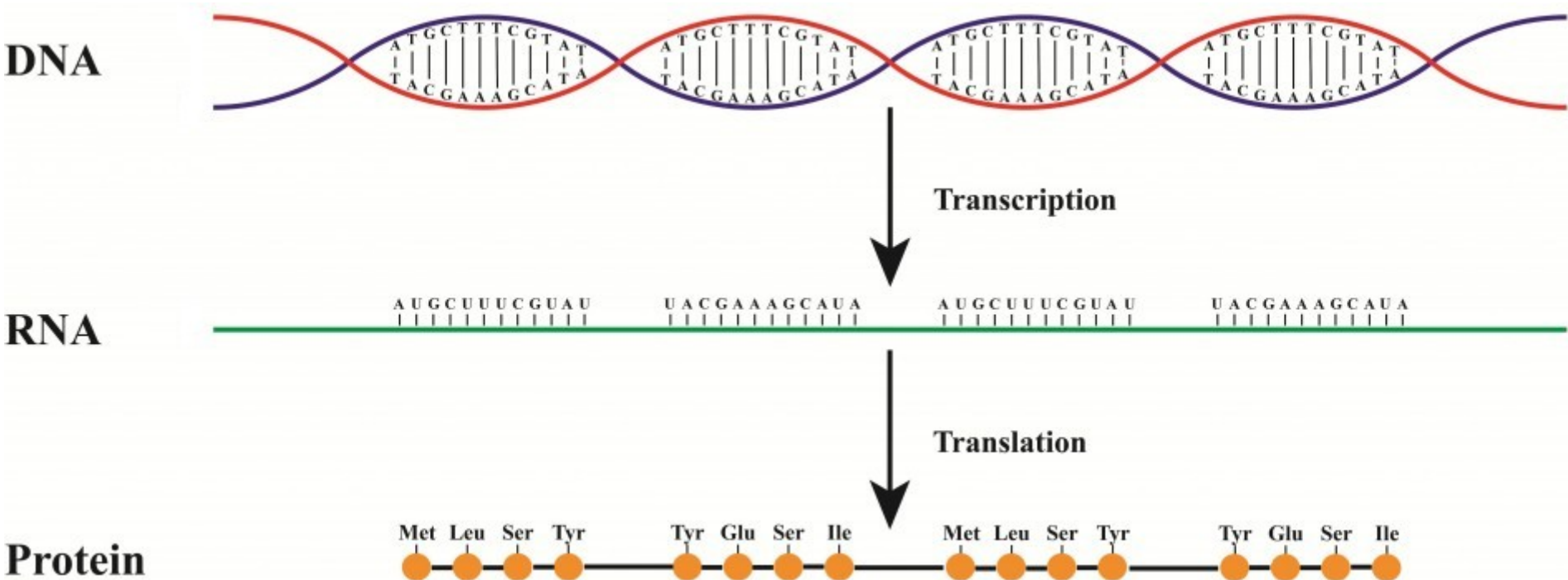
- Los tres grupos son unidos por un átomo de carbono llamado α -Carbono o C_{α} . A continuación vemos los 20 tipos de aminoácidos:

Nombre	Nombre inglés	Código una letra	Código tres letras
Valina	Valine	V	Val
Leucina	Leucine	L	Leu
Treonina	Threonine	T	Thr
Lisina	Lysine	K	Lys
Triptófano	Tryptophan	W	Trp
Histidina	Histidine	H	His
Fenilalanina	Phenylalanine	F	Phe
Isoleucina	Isoleucine	I	Ile
Arginina	Arginine	R	Arg
Metionina	Methionine	M	Met
Alanina	Alanine	A	Ala
Prolina	Proline	P	Pro
Glicina	Glycine	G	Gly
Serina	Serine	S	Ser
Cisteína	Cysteine	C	Cys
Asparagina	Asparagine	N	Asn
Glutamina	Glutamine	Q	Gln
Tirosina	Tyrosine	Y	Tyr
Ácido Aspártico	Aspartic Acid	D	Asp
Ácido Glutámico	Glutamic Acid	E	Glu

Proteína - Aminoácidos

- La información genética es traducida por la maquinaria celular para producir las proteínas usando el código genético, el cual determina la secuencia de aminoácidos codificada en el ADN y luego en el ARN.
- Durante la traducción la maquinaria celular utiliza la molécula de ARN como molde para sintetizar una cadena de aminoácidos codificada en la misma. Para ello interpreta el código leyendo de a 3 nucleótidos, esta unidad se denomina codón, cada codón codifica para un aminoácido.

Proteína - Aminoácidos



Proteína - Aminoácidos

		Segunda base					
		U	C	A	G		
P r i m e r a b a s e	U	Phe UUU	Ser UCU	Tyr UAU	Cys UGU	U	T e r c e r a b a s e
		Phe UUC	Ser UCC	Tyr UAC	Cys UGC	C	
		Leu UUA	Ser UCA	Stop UAA	Stop UGA	A	
		Leu UUG	Ser UCG	Stop UAG	Trp UGG	G	
	C	Leu CUU	Pro CCU	His CAU	Arg CGU	U	
		Leu CUC	Pro CCC	His CAC	Arg CGC	C	
		Leu CUA	Pro CCA	Gln CAA	Arg CGA	A	
		Leu CUG	Pro CCG	Gln CAG	Arg CGG	G	
	A	Ile AUU	Thr ACU	Asn AAU	Ser AGU	U	
		Ile AUC	Thr ACC	Asn AAC	Ser AGC	C	
		Ile AUA	Thr ACA	Lys AAA	Arg AGA	A	
		Met AUG	Thr ACG	Lys AAG	Arg AGG	G	
	G	Val GUU	Ala GCU	Asp GAU	Gly GGU	U	
		Val GUC	Ala GCC	Asp GAC	Gly GGC	C	
		Val GUA	Ala GCA	Glu GAA	Gly GGA	A	
		Val GUG	Ala GCG	Glu GAG	Gly GGG	G	

Nombre	Nombre inglés	Código una letra	Código tres letras
Valina	Valine	V	Val
Leucina	Leucine	L	Leu
Treonina	Threonine	T	Thr
Lisina	Lysine	K	Lys
Triptófano	Tryptophan	W	Trp
Histidina	Histidine	H	His
Fenilalanina	Phenylalanine	F	Phe
Isoleucina	Isoleucine	I	Ile
Arginina	Arginine	R	Arg
Metionina	Methionine	M	Met
Alanina	Alanine	A	Ala
Prolina	Proline	P	Pro
Glicina	Glycine	G	Gly
Serina	Serine	S	Ser
Cisteína	Cysteine	C	Cys
Asparagina	Asparagine	N	Asn
Glutamina	Glutamine	Q	Gln
Tirosina	Tyrosine	Y	Tyr
Ácido Aspártico	Aspartic Acid	D	Asp
Ácido Glutámico	Glutamic Acid	E	Glu

ACA-GAC-AGA-UAC-AAU

se traduce a

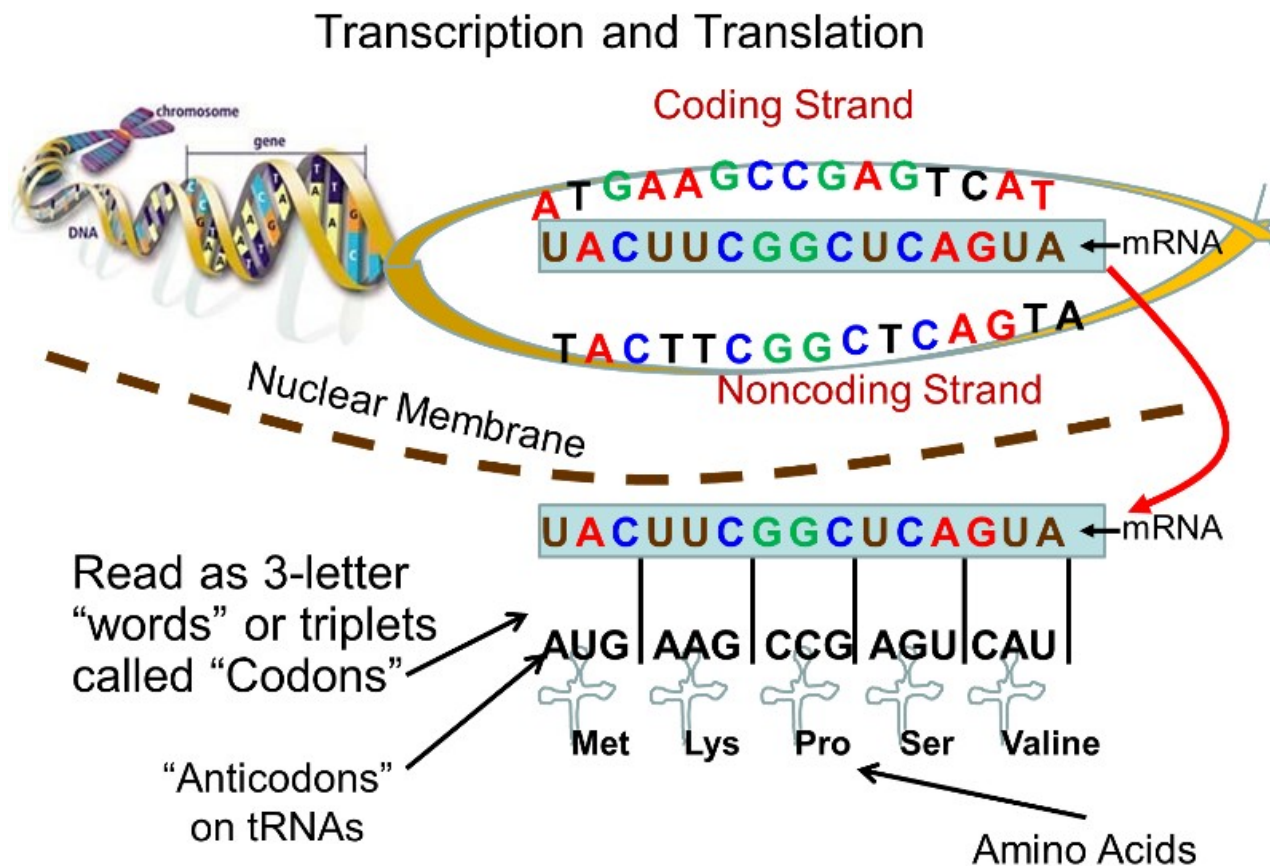
T D R Y N

Proteína - Aminoácidos

- Existen 64 combinaciones (codones) posibles formados a partir de combinar los 3 nucleótidos del ARN, sin embargo existen sólo 20 aminoácidos. Algunos aminoácidos están codificados en más de una combinación, además existen 3 combinaciones denominados stop. Cuando la maquinaria celular lee alguno de estas 3 combinaciones stop, la síntesis de la proteína codificada se detiene. Es decir, los codones stop determinan el final de la proteína y no codifican para ningún aminoácido (Lo codificaremos con un punto .).

Práctica 1 – 4 puntos

- Realice a mano y traduzca el siguiente diagrama:



Práctica 1 – 7 puntos

- Cree un programa que lea dos archivos practica1_1.txt y practica1_2.txt que lea una secuencia de ADN, realice la transcripción, la traducción y determine los aminoácidos en la cadena.
 - En caso de STOP considere un punto.
 - En caso exista una salto de línea la final, elimínelo.
 - Cada presentación son 2 puntos menos.
 - Cualquier lenguaje de programación.
 - Explicar el código.
 - Dos código iguales, se considerará plagio.
 - No se permite usar USB, se considerará plagio.

Práctica 1 – 7 puntos

		Segunda base					
		U	C	A	G		
P r i m e r a b a s e	U	Phe UUU	Ser UCU	Tyr UAU	Cys UGU	U	T e r c e r a b a s e
		Phe UUC	Ser UCC	Tyr UAC	Cys UGC	C	
		Leu UUA	Ser UCA	Stop UAA	Stop UGA	A	
		Leu UUG	Ser UCG	Stop UAG	Trp UGG	G	
	C	Leu CUU	Pro CCU	His CAU	Arg CGU	U	
		Leu CUC	Pro CCC	His CAC	Arg CGC	C	
		Leu CUA	Pro CCA	Gln CAA	Arg CGA	A	
		Leu CUG	Pro CCG	Gln CAG	Arg CGG	G	
	A	Ile AUU	Thr ACU	Asn AAU	Ser AGU	U	
		Ile AUC	Thr ACC	Asn AAC	Ser AGC	C	
		Ile AUA	Thr ACA	Lys AAA	Arg AGA	A	
		Met AUG	Thr ACG	Lys AAG	Arg AGG	G	
	G	Val GUU	Ala GCU	Asp GAU	Gly GGU	U	
		Val GUC	Ala GCC	Asp GAC	Gly GGC	C	
		Val GUA	Ala GCA	Glu GAA	Gly GGA	A	
		Val GUG	Ala GCG	Glu GAG	Gly GGG	G	

Nombre	Nombre inglés	Código una letra	Código tres letras
Valina	Valine	V	Val
Leucina	Leucine	L	Leu
Treonina	Threonine	T	Thr
Lisina	Lysine	K	Lys
Triptófano	Tryptophan	W	Trp
Histidina	Histidine	H	His
Fenilalanina	Phenylalanine	F	Phe
Isoleucina	Isoleucine	I	Ile
Arginina	Arginine	R	Arg
Metionina	Methionine	M	Met
Alanina	Alanine	A	Ala
Prolina	Proline	P	Pro
Glicina	Glycine	G	Gly
Serina	Serine	S	Ser
Cisteína	Cysteine	C	Cys
Asparagina	Asparagine	N	Asn
Glutamina	Glutamine	Q	Gln
Tirosina	Tyrosine	Y	Tyr
Ácido Aspártico	Aspartic Acid	D	Asp
Ácido Glutámico	Glutamic Acid	E	Glu

Práctica 1 – 7 puntos

- Muestre los resultados al profesor.
- Envíe el código fuente y los resultados al profesor vía correo electrónico (ehinojosa@unsa.edu.pe):
 - Asunto: Práctica 1 – CMB – UNSA - ApellidosNombres

Práctica 1 – 9 puntos

- Cree un programa que lea el archivo descargado de la Protein Data Bank (PDB) (Banco de Datos de Proteínas), para las siguientes proteínas (FASTA Sequence):
 - Orexina (PDB: 1CQ0)
 - Neuropeptido (PDB: 1RON)
 - Actina (PDB: 1J6Z)
 - Miosina (PDB 2V26)

Práctica 1 – 9 puntos

- Que realice el proceso inverso a la transcripción y la traducción y determine el ARN mensajero.
 - Cada presentación son 2 puntos menos.
 - Cualquier lenguaje de programación.
 - Explicar el código.
 - Dos código iguales, se considerará plagio.
 - No se permite usar USB, se considerará plagio.

Práctica 1 – 9 puntos

RCSB PDB Deposit ▾ Search ▾ Visualize ▾ Analyze ▾ Download ▾ Learn ▾ More ▾ MyPDB Login ▾

PDB-101 WORLDWIDE PDB PROTEIN DATA BANK EMDatabank Certified Data Resources for NMR ndb NUCLEIC ACID DATABASE StructuralBiology Knowledgebase Worldwide Protein Data Bank Foundation

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment

NMR Ensemble

1CQ0

SOLUTION STRUCTURE OF A HUMAN HYPOCRETIN-2/OREXIN-B 'STRUCTURE OF A HUMAN HYPOCRETIN-2/OREXIN-B '

DOI: 10.2210/pdb1cq0/pdb

Classification: [DE NOVO PROTEIN](#)

Display Files ▾ Download Files ▾

- FASTA Sequence
- PDB Format
- PDB Format (gz)

Práctica 1 – 9 puntos

		Segunda base					
		U	C	A	G		
P r i m e r a b a s e	U	Phe UUU	Ser UCU	Tyr UAU	Cys UGU	U	T e r c e r a b a s e
		Phe UUC	Ser UCC	Tyr UAC	Cys UGC	C	
		Leu UUA	Ser UCA	Stop UAA	Stop UGA	A	
		Leu UUG	Ser UCG	Stop UAG	Trp UGG	G	
	C	Leu CUU	Pro CCU	His CAU	Arg CGU	U	
		Leu CUC	Pro CCC	His CAC	Arg CGC	C	
		Leu CUA	Pro CCA	Gln CAA	Arg CGA	A	
		Leu CUG	Pro CCG	Gln CAG	Arg CGG	G	
	A	Ile AUU	Thr ACU	Asn AAU	Ser AGU	U	
		Ile AUC	Thr ACC	Asn AAC	Ser AGC	C	
		Ile AUA	Thr ACA	Lys AAA	Arg AGA	A	
		Met AUG	Thr ACG	Lys AAG	Arg AGG	G	
	G	Val GUU	Ala GCU	Asp GAU	Gly GGU	U	
		Val GUC	Ala GCC	Asp GAC	Gly GGC	C	
		Val GUA	Ala GCA	Glu GAA	Gly GGA	A	
		Val GUG	Ala GCG	Glu GAG	Gly GGG	G	

Nombre	Nombre inglés	Código una letra	Código tres letras
Valina	Valine	V	Val
Leucina	Leucine	L	Leu
Treonina	Threonine	T	Thr
Lisina	Lysine	K	Lys
Triptófano	Tryptophan	W	Trp
Histidina	Histidine	H	His
Fenilalanina	Phenylalanine	F	Phe
Isoleucina	Isoleucine	I	Ile
Arginina	Arginine	R	Arg
Metionina	Methionine	M	Met
Alanina	Alanine	A	Ala
Prolina	Proline	P	Pro
Glicina	Glycine	G	Gly
Serina	Serine	S	Ser
Cisteína	Cysteine	C	Cys
Asparagina	Asparagine	N	Asn
Glutamina	Glutamine	Q	Gln
Tirosina	Tyrosine	Y	Tyr
Ácido Aspártico	Aspartic Acid	D	Asp
Ácido Glutámico	Glutamic Acid	E	Glu

Práctica 1 – 9 puntos

- Muestre los resultados al profesor.
- Envíe el código fuente y los resultados al profesor vía correo electrónico (ehinojosa@unsa.edu.pe) (El mismo correo anterior):
 - Asunto: Práctica 1 – CMB – UNSA - ApellidosNombres

Gracias

Dr. Edward Hinojosa Cárdenas
ehinojosa@unsa.edu.pe